Some problems in orthogonal distance and non-orthogonal distance regression

G. A. Watson

Department of Mathematics, University of Dundee, Dundee DD1 4HN, Scotland.
gawatson@maths.dundee.ac.uk

Abstract

Of interest here is the problem of fitting a curve or surface to given data by minimizing some norm of the distances from the points to the surface. These distances may be measured orthogonally to the surface, giving orthogonal distance regression, and for this problem, the least squares norm has attracted most attention. Here we will look at two other important criteria, the l_1 norm and the Chebyshev norm. The former is of value when the data contain wild points, the latter in the context of accept/reject criteria. There are however circumstances when it is not appropriate to force the distances to be orthogonal, and two possibilities of this are also considered. The first arises when the distances are aligned with certain fixed directions, and the second when angular information is available about the measured data points. For the least squares norm, we will consider some algorithmic developments for these problems.

1 Introduction

Of interest here is the problem of fitting to given data a curve or surface which depends on a vector $\mathbf{a} \in R^n$ of parameters. The underlying approach is such that (1) a point on the surface is associated with each data point, (2) the fit of the surface is measured by a norm of the vector whose components are the distances between each pair of corresponding points, (3) the (correct) Gauss-Newton steps in $\mathbf{a}$ are used as a basis for minimizing this norm. The distances may be orthogonal to the surface, giving orthogonal distance regression (ODR), or may be forced to satisfy some other criterion which makes them non-orthogonal in general. We consider both situations.

For the ODR problem, most attention has been given to the least squares norm (eg [5], [8], [9], [16], [17], [22]). Here we will look at two other important criteria, the l_1 norm and the Chebyshev norm. The former is of value when the data contain wild points, the latter in the context of accept/reject criteria. For the non-orthogonal distance problem we will restrict attention to the least squares case.

In terms of a vector $\mathbf{a} \in R^n$ of parameters, the curve or surface may be defined in two ways, (a) **parametrically,** when a point $\mathbf{x}$ on the surface is given by

$$\mathbf{x} = \mathbf{x}(\mathbf{a}, \mathbf{t}),$$

with $\mathbf{t}$ the parameters whose values define the particular point, or (b) **implicitly**, when the surface is defined by the set of points $\mathbf{x}$ satisfying the scalar equation

$$f(\mathbf{a}, \mathbf{x}) = 0.$$

It is also assumed here that the expressions required in these representations are differentiable functions of their parameters.

2 l_1 and l_∞ ODR

Consider first the l_1 case. Then the problem is

$$\text{minimize} \sum_{i=1}^{m} \|\mathbf{x}_i - \mathbf{z}_i(\mathbf{a})\|,$$

where the points $\mathbf{z}_i(\mathbf{a})$ are the nearest points to $\mathbf{x}_i$ on the surface defined by $\mathbf{a}$, and where we will assume throughout that unadorned norms are Euclidean norms. Let

$$\delta_i = \|\mathbf{x}_i - \mathbf{z}_i(\mathbf{a})\|, \ i = 1, \ldots, m.$$

Then the problem is effectively now defined in terms of the vector $\mathbf{a}$ alone. It is easy to to calculate the correct Gauss-Newton step in $\mathbf{a}$, which minimizes

$$\|\delta + \nabla_{\mathbf{a}} \delta \mathbf{d}\|_1$$

with respect to $\mathbf{d}$. Now

$$\nabla_{\mathbf{a}} \delta_i = -\frac{(\mathbf{x}_i - \mathbf{z}_i(\mathbf{a}))^T}{\delta_i} \nabla_{\mathbf{a}} z_i(\mathbf{a}), \ \delta_i \neq 0,$$

so that there are potential problems if any $\delta_i \to 0$. Given the nature of the l_1 problem, we cannot exclude that possibility. In fact although δ is not a smooth function, because derivative discontinuities only occur at zero values it is a **strong semi-smooth function**, as defined in [12]. Ideas from smooth analysis and from strong semi-smooth analysis as developed in [11] can then be combined to give a local convergence analysis for the present problem. Fast local convergence for the usual smooth problem relies on strong uniqueness [4]; for the l_1 norm, this can be interpreted in terms of a requirement that the sequence of solutions $\mathbf{d}^k$ is "well-behaved" in a certain sense [1]. An analogous requirement can be stated here.

Let the current approximation be $\mathbf{a}^k$ and let J^k denote the Jacobian matrix $\nabla_{\mathbf{a}} \delta(\mathbf{a}^k)$, assuming this exists. Then the Gauss-Newton step $\mathbf{d}^k$ minimzes

$$\|\delta(\mathbf{a}^k) + J^k \mathbf{d}\|_1.$$

It is well known (see for example [18]) that if J^k has full rank then there always exists a solution $\mathbf{d}^k$ and an index set Z^k containing n indices such that

$$\delta_i(\mathbf{a}^k) + \mathbf{e}_i^T J^k \mathbf{d}^k = 0, \ i \in Z^k,$$

where $\mathbf{e}_i$ is the ith coordinate vector. Let $\mathbf{a}^*$ be a limit point of the iteration. Then for $\mathbf{a}^k$ close enough to $\mathbf{a}^*$, assume that J^k exists and

(i) $\delta(\mathbf{a}^k) + J^k \mathbf{d}^k$ has **exactly** n zeros, corresponding to an index set Z^k,

(ii) $Z^k = Z^*$, independent of k,

(iii) the $n \times n$ matrices whose rows are $\mathbf{e}_i^T J^k, i \in Z^*$, are bounded away from singularity.

In practice these conditions ensure that $\mathbf{d}^k$ is unique, and there is no redundancy in the zero components. An analysis is given in [21] for both parametric and implicit fitting. The main result is the following.

Theorem 2.1 *[21] Let the Gauss-Newton method produce a sequence $\mathbf{a}^k \to \mathbf{a}^*$, where $\delta(\mathbf{a}^k)$ has no zero components, and let (i)–(iii) above hold. In the parametric case, assume that for all $i \in Z^*$, there exists a unique unit normal vector $\mathbf{n}_i$ (up to change of sign) at the point $\mathbf{x}_i$ on the surface defined by $\mathbf{a}^*$. Then the (undamped) Gauss-Newton method converges to $\mathbf{a}^*$ at a second order rate.*

The significance of this result is that, for both parametric and implicit fitting, any δ_i tending to zero is not **by itself** necessarily an obstacle to good performance of the Gauss-Newton method in the l_1 case. What is more significant is the possibility of very slow convergence and this has more to do with the **number** of those zero components of δ at a limit point, rather than just their presence. A fundamental requirement for the condition (ii) is that the number of zero components of $\delta(\mathbf{a}^*)$ is n. Of course, this condition is a rather special one, and for many problems, will not be satisfied. There is slow (possibly very slow) convergence associated with this case.

Turning now to the l_∞ problem, this can be stated

$$\text{minimize } \max_i \|\mathbf{x}_i - \mathbf{z}_i(\mathbf{a})\|,$$

with $\mathbf{z}_i(\mathbf{a})$ defined as before. Again $\delta_i = \|\mathbf{x}_i - \mathbf{z}_i(\mathbf{a})\|$ is not a smooth function, but a solution normally occurs in a region where δ is smooth. Therefore the problem does not differ significantly from the usual nonlinear minimax problem: the main requirement for fast local convergence is that at a limit point the norm is attained at $n+1$ indices [4].

Two simple examples in 2 dimensions are given by way of illustration. A standard line search is incorporated to force global convergence, although trust region methods are a popular alternative. Indeed, local convergence is the main concern here, and we have not begun to address important issues to do with the development of robust general purpose algorithms.

Example 2.2 Consider the Späth data set [13] ($m = 7$), and consider fitting an ellipse defined implicitly, using the l_∞ and l_1 norms. The solutions are illustrated in Figure 1, where the dashed ellipse and dashed lines are the l_∞ solution and corresponding orthogonal directions, and the solid ellipse and solid lines are the l_1 solution and corresponding directions. Both ellipses were obtained using the Gauss-Newton method starting from the circle centre (5,5), radius 2, in 4 and 5 iterations respectively for 5 figure accuracy.

Example 2.3 Consider next the GGS data set [6], which has $m = 8$. Similar fits to those for Example 1 are shown in Figure 2. Again the Gauss-Newton method was used starting from the circle centre (5,5), radius 2, to give convergence in 6 iterations (l_∞) and 7 iterations (l_1).

For both these examples $n = 5$, and favourable conditions hold so that there is quadratic convergence both in the l_1 and l_∞ cases. Otherwise, the key to recovering fast

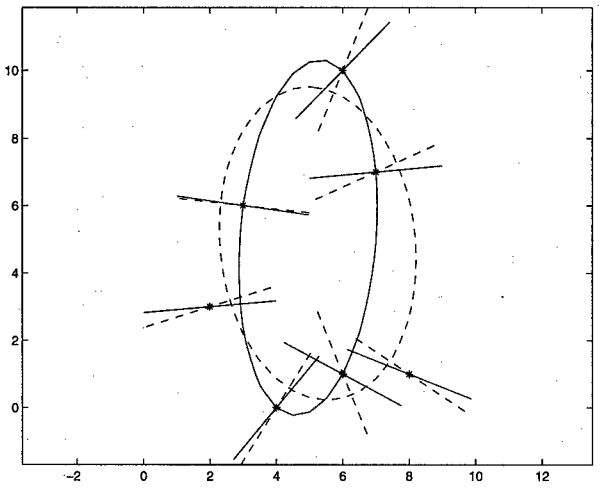

FIG. 1. l_1 and l_∞ fits to Späth data set.

local convergence in the l_1 case is to identify Z^* and to reformulate the problem locally as

$$\text{minimize} \sum_{i \notin Z^*} \|\mathbf{x}_i - \mathbf{z}_i(\mathbf{a})\| \text{ subject to } \mathbf{x}_i - \mathbf{z}_i(\mathbf{a}) = 0,\ i \in Z^*. \tag{2.1}$$

A similar remedy in the l_∞ case is as follows. For a limit point $\mathbf{a}^*$ of the iteration, let

$$I^* = \{i : \delta_i(\mathbf{a}^*) = \max_i \delta_i(\mathbf{a}^*)\}.$$

Then if we can identify I^*, $\mathbf{a}^*$ solves, for any $j \in I^*$:

$$\text{minimize } \delta_j(\mathbf{a}) \text{ subject to } \delta_i(\mathbf{a}) - \delta_j(\mathbf{a}) = 0,\ i \in I^* \backslash j.$$

Example 2.4 Fitting an l_∞ ODR line in R^3 to 100 random data points (equivalent to finding the circumscribing cylinder of smallest radius) gives slow convergence of the basic method, because $|I^*| = 3$ and $n = 4$. But once we identify $I^* = \{4, 42, 58\}$, only 5 iterations of the NAG Fortran subroutine E04UCF are required for 6 figure accuracy.

3 Non-orthogonal l_2 distance regression

3.1 Using fixed directions

Suppose that the data come from sampling the surface of a manufactured part, using a coordinate measuring machine with a touch probe. It has been argued by Hulting [10] that choosing the directions to be the known probe directions $\mathbf{v}_i$ (relative to a fixed frame of reference) not only makes explicit use of the measurement design, but

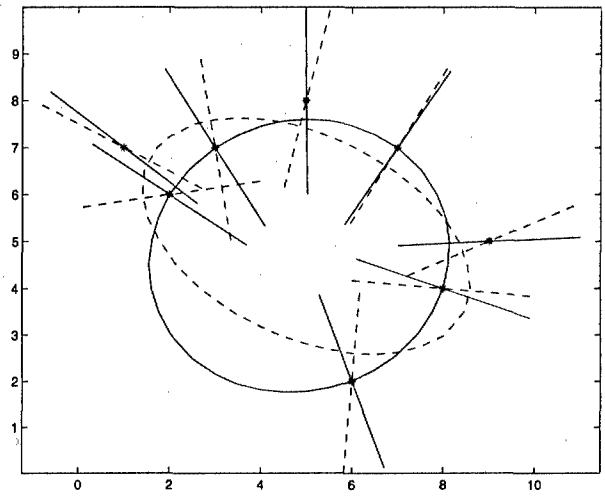

FIG. 2. l_1 and l_∞ fits to GGS data set.

also complies with traditional fixed-regressor assumptions (enabling standard inference theory to apply).

Let $\mathbf{x}_i$, $i = 1, \ldots, m$ as usual be the data points, and let $\mathbf{z}_i$ be the corresponding points on the surface reached by travelling along the lines from $\mathbf{x}_i$ in the direction $\mathbf{v}_i$. Then we require to minimize $\|\delta\|$ where

$$\delta_i = \|\mathbf{x}_i - \mathbf{z}_i(\mathbf{a})\|, \ i = 1, \ldots, m,$$

with $\mathbf{z}_i(\mathbf{a})$ defined by

$$\mathbf{z}_i(\mathbf{a}) - \mathbf{x}_i = \delta_i \mathbf{v}_i, \ i = 1, \ldots, m,$$

where $\mathbf{v}_i$ satisfying $\mathbf{v}_i^T \mathbf{v}_i = 1$ is given for each i. In case of ambiguity, the smallest value of δ_i is chosen. The basic idea in efficient algorithmic development is again to treat the problem as one in $\mathbf{a}$ alone, which can be solved as before by the Gauss-Newton method (or variants). Let $\mathbf{a}$ be given. Then for each point $\mathbf{x}_i$, the point where the line through $\mathbf{x}_i$ in the direction $\mathbf{v}_i$ first cuts the surface can be obtained (this calculation replaces the "footpoint problem" of calculating $\mathbf{z}_i(\mathbf{a})$ as the point on the surface in the orthogonal distance problem), giving δ_i as a function of $\mathbf{a}$. Methods based on Gauss-Newton steps are developed for the parametric case in [19], [20], and for the implicit case in [7].

By way of illustration, the 2 data sets previously considered in Examples 1 and 2 are used to fit ellipses defined implicitly with a particular choice of directions $\mathbf{v}_i$. The initial (circles) and final ellipses (together with the data points and the directions $\mathbf{v}_i$) are shown in Figures 3 and 4. The calculations needed respectively 19 and 17 iterations, reflecting the fact that, unlike the l_1 and l_∞ cases, the convergence rate is linear.

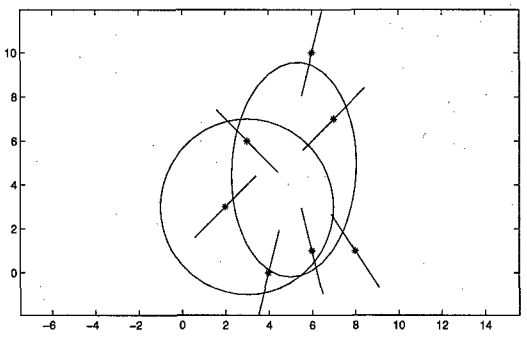

FIG. 3. l_2 fit to Späth data set: fixed $\mathbf{v}_i$.

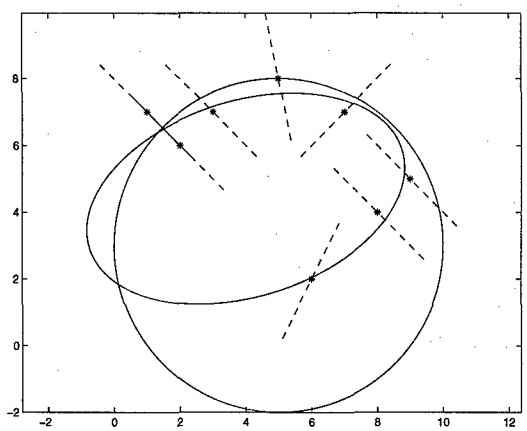

FIG. 4. l_2 fit to GGS data set: fixed $\mathbf{v}_i$.

3.2 Using angular information

Berman and Griffiths [2, 3] consider fitting a circle when angular differences between successively measured data points are known, with applications in physics and archaeology. This fitting problem has been extended to the case of ellipses and ellipsoids by Späth in [14, 15] and it is this kind of problem which is of interest here. The methods of [14] and [15] are based on the alternating algorithm, and while this can be perhaps surprisingly effective (particularly with a reparameterization of the problem), we consider here a correct separated Gauss-Newton method similar to that used before. In addition to (usually) better local convergence properties, standard step-length control can be incorporated.

To illustrate, consider fitting an ellipse in general position. It is convenient to do this by allowing the data to rotate, and fitting to those an ellipse in normal position, aligned with the axes. Let (x, y) denote the components of $\mathbf{x}$. Then we work with the data

$$x_i(\phi) = x_i \cos\phi + y_i \sin\phi, \;\; y_i(\phi) = -x_i \sin\phi + y_i \cos\phi,$$

for $i = 1, \ldots, m$, where ϕ is an unknown parameter. Therefore we require to minimize, with respect to the 6 parameters a, b, p, q, α, ϕ, the function

$$\sum_{i=1}^{m} \{(x_i(\phi) - a - p\cos(\alpha + t_i))^2 + (y_i(\phi) - b - q\sin(\alpha + t_i))^2\},$$

where the numbers t_i are given. Because $(\alpha + t_{i+1}) - (\alpha + t_i) = t_{i+1} - t_i$, for each i, we can interpret this as saying say that the angular differences are known, with a degree of freedom given by the parameter α. Note that at a solution to this problem, the directions between pairs of points $(x_i(\phi), y_i(\phi))$ and the corresponding points on the ellipse will not generally be orthogonal to the ellipse.

Differentiating the above expression with respect to a, p, b, q gives

$$A_1 \begin{bmatrix} a \\ p \end{bmatrix} = \mathbf{c}_1, \tag{3.1}$$

where

$$A_1 = \begin{bmatrix} m & \sum_{i=1}^m \cos(\alpha + t_i) \\ \sum_{i=1}^m \cos(\alpha + t_i) & \sum_{i=1}^m \cos^2(\alpha + t_i) \end{bmatrix}, \mathbf{c}_1 = \begin{bmatrix} \sum_{i=1}^m x_i(\phi) \\ \sum_{i=1}^m x_i(\phi)\cos(\alpha + t_i) \end{bmatrix};$$

$$A_2 \begin{bmatrix} b \\ q \end{bmatrix} = \mathbf{c}_2, \tag{3.2}$$

where

$$A_2 = \begin{bmatrix} m & \sum_{i=1}^m \sin(\alpha + t_i) \\ \sum_{i=1}^m \sin(\alpha + t_i) & \sum_{i=1}^m \sin^2(\alpha + t_i) \end{bmatrix}, \mathbf{c}_2 = \begin{bmatrix} \sum_{i=1}^m y_i(\phi) \\ \sum_{i=1}^m y_i(\phi)\sin(\alpha + t_i) \end{bmatrix}.$$

Then (3.1) and (3.2) give (a, b, p, q) as functions of α and ϕ, provided that A_1 and A_2 are nonsingular: this will be assumed. For given α and ϕ, we can therefore define the function to be minimized as

$$F(\alpha, \phi) = \|\delta(\alpha, \phi)\|,$$

where

$$\delta_i = \|\mathbf{w}_i\|, \;\; i = 1, \ldots, m, \tag{3.3}$$

with

$$\mathbf{w}_i = (x_i(\phi) - a - p\cos(\alpha + t_i), \;\; y_i(\phi) - b - q\sin(\alpha + t_i))^T,$$

and with a, b, p, q defined by (3.1) and (3.2). Then we can apply the Gauss-Newton method to the minimization of $F(\alpha, \phi)$. The basic step $\mathbf{d} = (\delta\alpha, \delta\phi)^T$ is given by finding

$$\min_{\mathbf{d} \in R^2} \|\delta + J\mathbf{d}\|, \tag{3.4}$$

where $J \in R^{m\times 2}$ has ith row given by

$$\mathbf{e}_i^T J = \nabla_{\alpha,\phi}\delta_i(\alpha,\phi),\ i = 1,\ldots,m.$$

Now

$$\nabla_{\alpha,\phi}\delta_i(\alpha,\phi) = \frac{\mathbf{w}_i^T}{\delta_i}(\nabla_{\alpha,\phi}\mathbf{w}_i + (\nabla_{a,p,b,q}\mathbf{w}_i)M),\ \delta_i \neq 0, \tag{3.5}$$

where

$$M = \nabla_{\alpha,\phi}\begin{pmatrix} a \\ p \\ b \\ q \end{pmatrix} \in R^{4\times 2}.$$

It is easy to compute M from (3.1) and (3.2) which can be interpreted as identities in α and ϕ. The details are omitted, but all the linear systems use just the matrices A_1 and A_2, and apart from the solution of (3.4) (a least squares problem in two variables), there remains only evaluation of expressions.

Example 3.1 Consider Example 1 from [14], which has $m = 11$. Starting from $\alpha = 0$, $\phi = 0$, 15 iterations are required to satisfy the stopping criterion $\|\mathbf{d}\|_\infty < 0.001$. The resulting value of $\|\delta\|^2$ is 7.7211, with $a = 2.1253$, $b = -0.1700$, $p = 4.1281$, $q = 3.0931$, $\alpha = 13.2348°$, $\phi = 34.7309°$.

Example 3.2 Next consider Example 2 from [14], which has $m = 8$. Again starting from $\alpha = 0$, $\phi = 0$, 9 iterations are required to satisfy the stopping criterion $\|\mathbf{d}\|_\infty < 0.001$. The resulting value of $\|\delta\|^2$ is 4.4946, with $a = 4.3608$, $b = 1.9537$, $p = 5.3717$, $q = 3.3704$, $\alpha = -0.6215°$, $\phi = 26.3889°$.

4 Conclusions

We have examined some aspects of fitting curves and surfaces to given data. The underlying criterion involves associating with each data point a point on the surface and minimizing some norm of the vector whose components are the distances between pairs of points. The distances can be orthogonal to the surface, or fixed in some other way. But the problems have in common that methods based on separated Gauss-Newton steps can readily be developed.

Bibliography

1. Anderson, D. H. and M. R. Osborne, Discrete, non-linear approximation problems in polyhedral norms, Num. Math. **28**, 143–156 (1977).
2. Berman, M., Estimating the parameters of a circle when angular differences are known, Appl. Statist. **32**, 1–6 (1983).
3. Berman, M. and D. Griffiths, Incorporating angular information into models for stone circle data, Appl. Statist. **34**, 237–245 (1985).
4. Cromme, L., Strong uniqueness: a far reaching criterion for the convergence of iterative processes, Numer. Math. **29**, 179–193 (1978)

5. Forbes, A. B., Least squares best fit geometric elements, in *Algorithms for Approximation II*, eds. J. C. Mason and M. G. Cox, Chapman and Hall, London, 311–319 (1990).
6. Gander, W., G. H. Golub and R. Strebel, Fitting of circles and ellipses: least square solution, BIT **34**, 556 –577 (1994).
7. Gulliksson, M., I. Söderkvist and G. A. Watson, Implicit surface fitting using directional constraints, BIT **41**, 331–344 (2001).
8. Helfrich, H.-P. and D. Zwick, A trust region method for implicit orthogonal distance regression, Numer. Alg. **5**, 535–545 (1993).
9. Helfrich, H.-P. and D. Zwick, A trust region algorithm for parametric curve and surface fitting, J. Comp. Appl. Math. **73**, 119–134 (1996).
10. Hulting, F. L., Discussion contribution to the paper by M. M. Dowling, P. M. Griffin, K.-L. Tsui and C. Zhou, Statistical issues in geometric feature inspection using coordinate measuring machines, Technometrics **39**, 18–20 (1997).
11. Qi, L., Convergence analysis of some algorithms for solving nonsmooth equations, Math. of Operations Research **18**, 227–244 (1993).
12. Qi, L. and G. Jiang, Semismooth Karush-Kuhn-Tucker equations and convergence analysis of Newton methods and quasi-Newton methods for solving these equations, Math. of Operations Research **22**, 301–325 (1997).
13. Späth, H., Least squares fitting by circles, Computing **57**, 179–185 (1996).
14. Späth, H., Estimating the parameters of an ellipse when angular differences are known, Comput. Stat. **14**, 491–500 (1999).
15. Späth, H. Least squares fitting of spheres and ellipsoids using not orthogonal distances, Math. Comm. **6**, 89–96 (2001).
16. Turner, D. A., *The Approximation of Cartesian Co-ordinate Data by Parametric Orthogonal Distance Regression*, PhD Thesis, University of Huddersfield (1999).
17. Turner, D. A., I. J. Anderson, J. C. Mason, M. G. Cox and A. B. Forbes, An efficient separation-of-variables approach to parametric orthogonal distance regression, in *Advanced Mathematical and Computational Tools in Metrology IV*, eds P. Ciarlini, A. B. Forbes, F. Pavese and D. Richter, Series on Advances in Mathematics for Applied Sciences, Volume 53, World Scientific, Singapore, 246–255 (2000).
18. Watson, G. A., *Approximation Theory and Numerical Methods*, John Wiley, Chichester (1980).
19. Watson, G. A., Least squares fitting of circles and ellipses to measured data, BIT **39**, 176–191 (1999).
20. Watson, G. A., Least squares fitting of parametric surfaces to measured data, ANZIAM J **42** (E), C68–C95 (2000).
21. Watson, G. A, On the Gauss-Newton method for l_1 orthogonal distance regression, IMAJ. Num. Anal. (to appear).
22. Zwick, D. S., Applications of orthogonal distance regression in metrology, in *Recent Advances in Total Least Squares and Errors-in-Variables Techniques*, ed S. Van Huffel, SIAM, Philadelphia, pp. 265–272 (1997).